Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 35 (2014) 598 – 605

Procedia Computer Science

18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems - KES2014

A smart experience-based knowledge analysis system (SEKAS)

Peng Wang*, Cesar Sanin and Edward Szczerbicki

School of Engineering, Faculty of Engineering and Built Environment
The University of Newcastle, Australia

Abstract

This article addresses the issues associated with using ever-increasing amounts of information and knowledge more effectively, and taking advantage of knowledge generated through experience.
A hybrid structure, the Smart Experience-based Knowledge Analysis System (SEKAS), is put forward in this paper to address issues of knowledge management and use. SEKAS combines a set of experience knowledge structures (SOEKS) with multiple techniques to provide a comprehensive knowledge management approach capturing, discovering, reusing and storing knowledge for the users. The SEKAS integrates a novel Decisional DNA (DDNA) knowledge structure with the traditional web crawler technologies. DDNA, as a knowledge representation platform, can help deal with noisy and incomplete data, with learning from experience, and with making precise decisions and predictions in vague and fuzzy environments.
The paper outlines the investigation of the combination of DDNA and feature selection algorithms to guarantee the future performance for prediction. The proposed approaches are general and extensible in terms of both designing novel algorithms, and in the application to other domains.
The SEKAS integrates the evolutionary algorithm, NSGA-II, using experience that is derived from a former decision event, to improve the evolutionary algorithm's ability to find optimal solutions rapidly and efficiently. The SEKAS application to solve a travelling salesman's problem shows that this new proposed hybrid model can find optimal or close to true, Pareto-optimal solutions in a fast and efficient way.

Peer-review under responsibility of KES International.

Keywords: Decisional DNA; Set of Experience Knowledge Structure; Knowledge Management; Data Mining, Evolutionary Algorithm

* Corresponding author.
E-mail address: peng.wang@uon.edu.au

Peer-review under responsibility of KES International.
doi:10.1016/j.procs.2014.08.141

1. Introduction

Knowledge management is of the utmost importance to organisations in terms of the integration, distribution and refinement of knowledge and information. This knowledge is only useful to organisations when used to solve problems and make decisions[4]. In order to support the knowledge management process, experience and knowledge have to be spread among individuals and organisations. Knowledge sharing not only fosters collaboration but also facilitates experience and knowledge discovery. Several methods have been developed to support collaboration and knowledge sharing using different technologies such as ontologies, the internet, and data repositories[2]. However, although the technologies improve the knowledge management, a comprehensive system that uses a domain independent knowledge representation, and has the ability to extract, compute and refine existing knowledge, remains the focus of a research area still to be explored.

The Smart Experienced-based Knowledge Analysis System (SEKAS) is proposed as a new technological model based on knowledge management and is designed to provide common approaches that allow different systems to capture, share, reuse, learn and evolve the experiential knowledge. SEKAS is based on the SOEKS model and presents knowledge as DDNA, combining different techniques to support intelligent and autonomous system in multiple domains. The motivation of this structure is to develop a comprehensive structure that combines different techniques to help organisations deal with their increasing information. The SEKAS supports the decision maker of the organisations to efficiently and accurately make decision in order to quickly adapt the requirement of markets. The SEKAS helps the companies share, reuse and manage knowledge, but also provide a smart tool to learn and evolve the captured knowledge for more complicated issues.

This paper depicts the general components proposed for the SEKAS, starting with its overall view and main features. Next, the conceptual models and architecture are presented, describing the different layers and services that are composed of the model. Finally, a summary and brief conclusion on this paper is presented. The aim of this paper is to provide an outline of the SEKAS.

2. Background

2.1. Set of experience knowledge structure (SOEKS) and decisional DNA

The Set of Experience Knowledge Structure (SOEKS or shortly SOE[4, 12] has been used to collect and store formal decisional events in an explicit manner[5].This structure as a flexible and independent knowledge representation is a suitable tool for capturing, reusing and sharing knowledge within organisations. Therefore, the SOEKS, that is based on existing and available knowledge offered by a formal decision event with dynamic structure, can be expressed in XML or OWL as ontology in order to make it shareable and transportable[6].

The SOEKS which is constituted by variables, functions, constraints and rules[7] is structured in view of some important features of DNA. Firstly, the combination of the four components of the SOEKS offers distinctiveness, just corresponding to the combination of the four nucleotides of DNA. The elements of the SOEKS imitate a gene to connect with each other by using those four elements. In the same way as a gene produces a phenotype, the SOEKS yields a value of decision with their elements. Each SOEKS can be categorised and acts as a gene in DNA[9]. A set of SOEKS in a same category makes up of a decisional chromosome which stores decisional strategies for that category. After this, each module of chromosomes establishes an entire inference tool to offer a blue print of knowledge inside an organization[7].

2.2. Data mining

Data mining is the process of automatically seeking large data repositories to discover patterns and trends that go beyond simple analysis[3]. Data mining uses sophisticated mathematical algorithms to classify the data and evaluate the probability of future events[14]. It allows users to study data from many different dimensions or angles in order to categorize it and summarize the relationship among data. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large data repositories.

Data mining is largely used today by companies with a strong consumer focus in many fields, such as, retail, financial, communication, and marketing organizations. It assists these companies to analyse relationships among internal factors such as price, product positioning, and external factors such as economic indicators, competition, and customer demographics. And, it enables the companies to have the ability to improve their business on sales, customer satisfaction, and corporate profits. Finally, it summarizes information to view detail transactional data.

2.3. Evolutionary algorithm

Evolutionary Algorithm (EA) is an adaptive heuristic search algorithm which was inspired by[1] that imitates the process of the natural selection. Commonly EA mimics the process of the nature such as selection, crossover, mutation and inheritance. EA can be used to solve the optimization problem. In addition, EA has been established as the appropriate measure, searching solutions for multi-objective problems that are too complex to be solved by exact methods.

The EA starts from a population which contains randomly generated individuals. It is an iterative process in which each process also called a generation. In each generation, the EA based on objective functions which can be used to measure optimization problem evaluates a fitness for every individual in the population. The individuals with better fitness are selected from current population, and then the individuals are through crossover, mutation and selection process to form a new generation. The new generation is then used in the next iteration. Normally, if a maximum number of generations has been produced or the fitness level are satisfied the requirement of the algorithm, the algorithm is terminated.

3. Overview of the SEKAS

Organisations are typically comprised of hundreds of applications that are custom-built, acquired from a third-party, part of a legacy system, or a combination of these, and operating in multiple tiers of different operating system platforms. Most applications in decision-making processes rely upon models and can solve only specific kinds of problems in certain domains. The SEKAS confronts problems in a more encompassing way, not confining the system to particular areas, but rather depending on the experiences it undertakes. The system is a smart, multi-source knowledge-based decision support system, leveraging an integrated model of rule-based systems, expert systems, numerical models, self-learning and intelligent technology to assist managers in the decision-making process. Consequently, the SEKAS is designed as a multi-functional structure, providing several measures to adapt different efficient knowledge discovery and utilisation using the latest advances in technology. The SEKAS is the way experience represented as SOEKS and DDNA is passed on, and evolves through generations of decision makers in an autonomous and smart fashion.

Fig. 1. depicts the overview of the SEKAS. Different organisations and individuals interact autonomously with each other through the SEKAS. The Internet and cloud computing are also accessed for the extension of knowledge acquisition by using the same system.

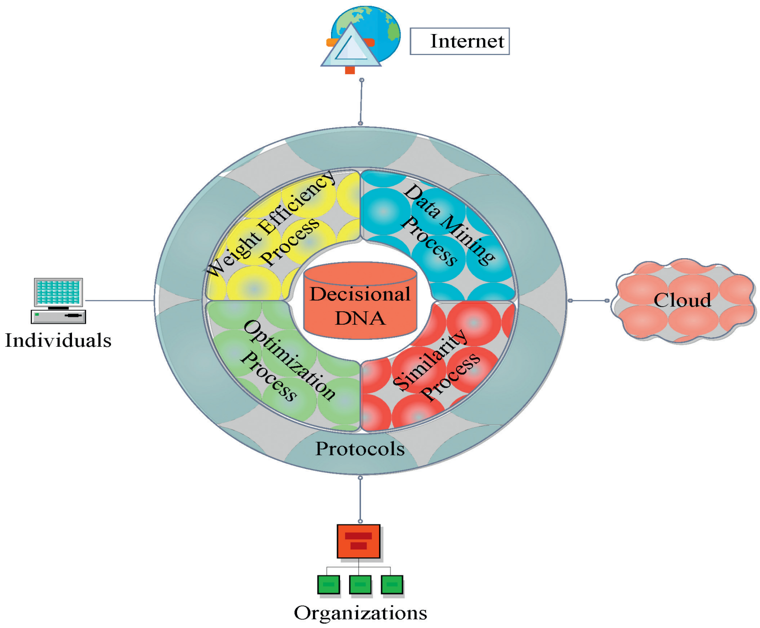

Fig. 1. The overview of the Smart Experienced-based Knowledge Analysis System.

The SEKAS has three major levels. At the outer level of the SEKAS, the protocols are created to interact with different organisations and individuals. These protocols translate and convert required information into a unique language and measurement model, thereby unifying information and making it understandable for the next level of the SEKAS. There are many advanced techniques, such as web crawlers, data adaptors and database adaptors, that are involved in this level in order to adapt to different information sources. The function level combines a variety of smart and efficient methods, such as data mining techniques and genetic algorithms. These tools are responsible for producing a set of proposed solutions or decisions so that then a single solution can be chosen according to the user's demands. The inner level is composed of the SOEKS and the DDNA. At this level, knowledge is continually renegotiated and classified by the issues that the knowledge belongs to.

4. Main features of SEKAS

In order to share experience and support decision-making processes in organisations, the SEKAS provides the following features:

- Experience-oriented: The model provides experience-oriented problem-solving capabilities. By using experience, decision makers can take advantage of modern computational improvements to support their organisation's decision-making processes, making decisions with increased efficiency and effectiveness.

- Applicability and transformability: The ability to access information in different systems among organisations, internally and externally, is the critical issues for knowledge management. For solving this problem, SEKAS is designed as an independent model based on the SOEKS and DDNA, being reusable by different applications within a given problem domain.

- Predictability and computational capability: Many researchers believe that their work will eventually be incorporated into a machine with general intelligence, to make the machine smarter and to replace human beings in inhospitable environments[14]. The ability to learn from experience and a set of solutions or predictions by calculation, is an important feature of the SEKAS. As a flexible knowledge representation structure, DDNA uses a set of symbols to represent a set of facts. It supplies a formal, semantic way to present how reasoning functions should be applied to the symbols in the model. The model integrates the DDNA with machine learning algorithms, such as the RELIEF-F algorithm[8] to enhance the systems that applied the SEKAS.

- Evolution capability: A core objective of the SEKAS is to perform accurately on new, unseen tasks, after having experienced a learning set of the SOEKS to generate new solutions from an existing experience. This functionality repeatedly mimics the process of natural selection inspired by biological evolution, such as reproduction, mutation, recombination and selection, by producing accurate decisions or solutions for decision makers. This feature is usually used to solve complex optimisation problems.

- Autonomous evolution of weight function: The SEKAS provides a mathematical tool to evaluate each element and assign weight to present the influence on the result. For instance, the system learns from existing experiences to find a pattern among them. Then, an evaluation method is employed to distribute the weight for each variable of the SOEKS, presenting the influence on the result variable. This process is constantly updating existing experiences with data and information from the real world, which is provided by the users, and the software applications they use.

5. Comprehensive structure of SEKAS

For carrying these five core features introduced above, the conceptual architecture is designed for SEKAS. It contains three layers: integration, operation and knowledge repository layers. The model is conceptualised on top of the knowledge management, to extend the capabilities of the latter by using DDNA and SOEKS for knowledge representation and exchange. Fig. 2. shows the conceptual architecture for the SEKAS.

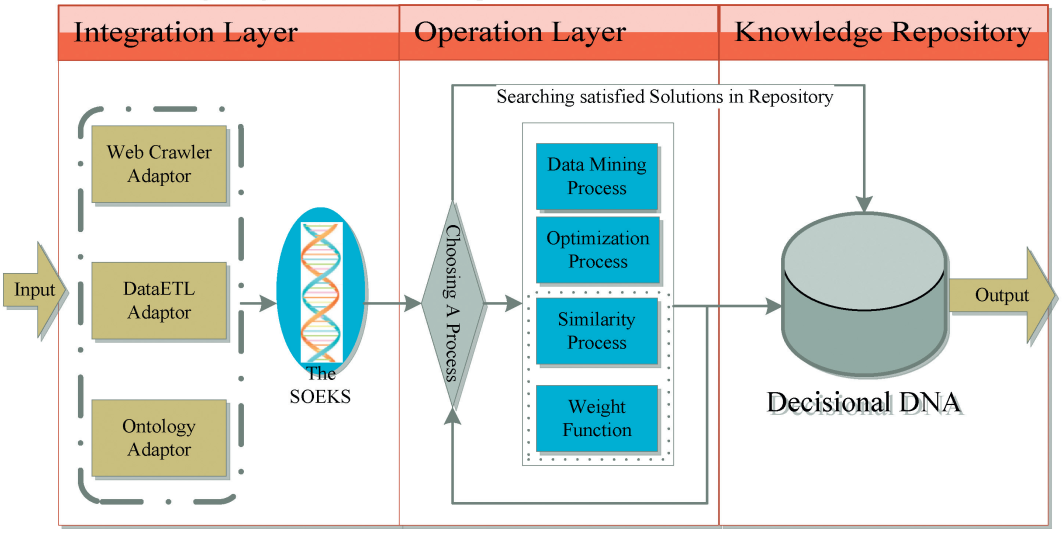

Fig. 2. Architecture of the Smart Experienced-based Knowledge Analysis Structure

The SEKAS is a tool that integrates different applications within and between organisations by sharing knowledge. This knowledge is acquired through an organisation's daily events and converted into the SOEKS and DDNA. Then, the gathered DDNA is operated by each individual process, depending on the goals that the decision-maker wants to achieve. Hence, this structure takes information and sends it through the three layers of integration, operation and knowledge repository.

5.1. The integration layer

The integration layer consolidates and integrates information offered by multiple applications. It identifies useful information and converts this data to a unique structure using the variables, functions, rules and constrains of the SOEKS. In other words, the integration layer assembles the characteristics and objectives of the organisation at a particular moment: it is a collector of frozen information and isolated solutions by using the SOEKS' capacities. Three adaptors - the web crawler, DataETL and ontology adaptors - comprise this layer in the system in order to adapt data with different structures.

- Web crawler adaptor

As one of the main sources of information, the internet plays an important role in daily life. Organisations have realised that acquiring knowledge from a website is an inevitable measure to sell more products, minimise costs and increase profits. In order to do this, the SEKAS focuses on marketing strategies and competitive analysis, or on the relationship with the customers. The different kinds of web data that are associated with customers are then categorised and gathered to build detailed customer profiles. This not only allows companies to retain current customers by being able to provide more personalised services, but also aids in attracting new customers.

Consequently the web crawler adaptor is implemented as an interface between the system and the website, collecting information and interpreting it with a unique language, such as the SOEKS or the DDNA, for further analysis. Because the required information still needs to be adjusted and manipulated, users still play an important role in this process. Users, based on their explicit objectives, collect data as pre-solutions or information, following the atomicity and multidisciplinary concepts.

- DataETL adaptor

The DataETL adaptor is designed by implementing the concept of extract, transform and load (ETL), which integrates data from multiple applications. It helps organisations to centralise data to increase transparency and efficiency, consolidate data for reduced maintenance costs, and increase query performance and usability. The first step of this adaptor is to extract data from source applications in which users must be involved to determine and identify the variables able to be studied, as well as other information attached to them such as values, functions, constraints and rules as the SOEKS structure required. Ultimately, the system loads the converted information as a chromosome into the end target for further process.

- Ontology adaptor

Ontology technologies are commonly used in artificial intelligence and knowledge representation. Computer programs can use ontologies for a variety of resolutions containing inductive reasoning, classification and problem-solving techniques, as well as communication and sharing of information among different systems. Furthermore, integrating semantic web systems use ontologies for a better interaction and understanding between different agent web-based systems[10]. DDNA and SOEKS can be represented using ontologies and OWL[6], exploiting powerful representation, querying and inference capabilities. Consequently, the adaptor interprets information from the OWL to ensure the model fundamentals are used.

5.2. The operation layer

Once the SOEKS or the DDNA is reached, the user can choose processes to find solutions or decisions. This is achieved by providing several models to calculate the solutions in the organisation, not only internally, but also externally. In addition, each model offers a scenario that produces measurements for the problems.

The multiple kinds of information produced by the applications are converted to variables and relationships, as well as many pre-decisions. Moreover, due to the fact that each application gives a pre-decision according to its established objectives, four processes provide measurements to seek appropriate solutions: data mining, optimisation, similarity and weight function processes.

- Data mining process

Data mining is a well-known and powerful technology, with great potential to assist companies to focus on the most important information in the data they have collected about the behaviour of their existing and potential customers. It discovers information within the data that queries and reports cannot effectively reveal. Consequently, using data mining techniques will enhance the knowledge management abilities of the SEKAS. This process is used to discover patterns and relationships in the data in order to help make better business decisions. It can help spot sales trends, develop smarter marketing campaigns, and accurately predict customer loyalty.

- Optimisation process

The optimisation process integrates the SOEKS and evolutionary algorithms to process a sequence of precisely defined steps that eventually lead to a solution to the given problem. It is one of the so-called optimisation methods, such as genetic algorithm, for searching optimums (global maximums or minima). This process can solve optimisation problems, such as multi-dimensional, non-differential, non-continuous and even non-parametrical problems, as these can be described with the SOEKS structured chromosome. The optimisation process that can be easily transferred to existing simulations and models explores new ways to put explicit knowledge in the hands of all internal and external stakeholders.

- Similarity and weight function processes

The similarity process adapts the discipline of multidimensional scaling to calculate the distance between different objects and is a useful tool to help an organisation in predictions, hypothesis testing and rule discovery[4, 11]. This technique assumes that data objects can be represented by values on continuous dimensions and that similarity can be presented by distance in a coordinate space. A problem with this approach is that the method is inappropriate for data objects that have a number of qualitative attributes[4, 12]. For this reason, a weight function has been proposed in order to evaluate the importance of each attribute that has influence on the results. The weight function learns the pattern from existing experience and then uses these patterns to assess the attributes and rank them in order of importance. Combining the similarity and weight function processes is an efficient way to establish an element of analysis among data objects, making them comparable and classifiable for helping in the decision-making process[4].

5.3. Knowledge repository layer

The knowledge repository layer is the last layer of the SEKAS and organises, categorises and ranks captured knowledge from the previous layer. Once the SOEKS has been originated by operating the processes in the operation layer, the knowledge repository layer begins. The layer interacts with the operation layer during the process of finding a solution by operating the processes that are chosen, and, as a result, the user finds a faster and more reliable solution. Two situations can occur within this repository:

- A SOEKS is interpreted and captured from applications with variables and values as a query. The system then searches the knowledge repository and identifies a group of suitable solutions ranked to the user. The user then decides whether the solution fits their requirements.

- User queries system with a SOE. If the system seeks the knowledge repository and cannot find the matched solution, the user then chooses a process in the operation layer for generating new solutions. This process will be repeatedly chosen until the optimal solution is identified.

This layer has the ability of auto-feeding. However, users must manage this layer due to the inherently dynamic nature of knowledge[5, 13]. The SEKAS provides suggestions when the system performs, but as decisions are concerned and taken in a different way by the user, those differences concerning the solutions should be re-evaluated automatically. Hence, the system undergoes a rising knowledge cycle based on built-in processes, as well as the decision-maker actions, in order to keep the system fresh and improved.

6. Conclusion and future work

In this paper, the notion of the SEKAS was introduced with an overview. The SEKAS is based on the principles of various computing technologies, such as knowledge management, data mining and heuristics.

Five features of this system were presented in the second section of this paper as the main functionalities in the work presented. Ultimately, this article provided a comprehensive structure of the SEKAS, and explained in detail three main layers and their purposes.

The future work will focus on the following:

Implementing an experiment based on data mining algorithms to learn the trends of data for predictions.

Evaluating the efficiency and effectiveness of the advanced evolutionary ability of the SEKAS by using famous optimization problems such as the travelling sales man.

References

1. Darwin, C. Perceiving vowels in the presence of another sound: A quantitative test of the "Old-plus-new" Heuristic. *Levels in Speech Communication: Relations and Interactions: A Tribute to Max WajskopAmsterdam, Elsevier;1995*, 1-12.
2. Kirkos, E., Spathis, C. & Manolopoulos, Y. Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications;2007,* 32**,** 995-1003.
3. Lavrač, N., Motoda, H., Fawcett, T., Holte, R., Langley, P. & Adriaans, P. Introduction: Lessons learned from data mining applications and collaborative problem solving. *Machine learning;2004,* 57**,** 13-34.
4. Sanin, C. 2007. *Smart Knowledge Management System.* PhD Thesis, Faculty of Engineering and Built Environment, School of Engineering, University of Newcastle, Newcastle, Australia.
5. Sanin, C. & Szczerbicki, E. Experience-based Knowledge Representation: SOEKS. *Cybernetics and Systems: An International Journal;2009,* 40**,** 99 - 122.
6. Sanin, C., Toro, C. & Szczerbicki, E. An OWL ontology of set of experience knowledge structure. *Journal of Universal Computer Science 13(2);2007***,** pp. 209–223.
7. Sanin, C., Toro, C., Zhang, H., Sanchez, E., Szczerbicki, E., Carrasco, E., Wang, P. & Mancilla-Amaya, L. Decisional DNA: A multi-technology shareable knowledge structure for decisional experience. *Neurocomputing;2012,* 88**,** 42-53.
8. Sun, Y. Iterative RELIEF for Feature Weighting: Algorithms, Theories, and Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence;2007,* 29.
9. Toro, C., Sanchez, E., Carrasco, E., Mancilla-Amaya, L., Sanín, C., Szczerbicki, E., Graña, M., Bonachela, P., Parra, C., Bueno, G. & Guijarro, F. Using Set Of Experience Knowledge Structure To Extend A Rule Set Of Clinical Decision Support System For Alzheimer's Disease Diagnosis. *Cybernetics and Systems;2012,* 43**,** 81-95.
10. W3c. 2004. *OWL Web Ontology Language Overview* [Online]. W3C Recommendation 10 February 2004. Available: http://www.w3.org/TR/2004/REC-owl-ref-20040210/ [Accessed 10 Novermber 2013].
11. Wang, P., Sanin, C. & Szczerbicki, E. Introducing The Concept of Decisional DNA-Based Web Content Mining. *Cybernetics and Systems: An International Journal;2012,* 43**,** 97-117.
12. Wang, P., Sanin, C. & Szczerbicki, E. Prediction Based on Integration of Decisional DNA and a Feature Selection Algorithm RELIEF-F. *Cybernetics and Systems;2013,* 44**,** 173-183.
13. Wenger, E. Knowledge management as a doughnut: Shaping your knowledge strategy through communities of practice. *Ivey Business Journal;2004,* 68**,** 1-8.
14. Witten, I. H. & Frank, E. 2005. *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann.